Feature by David Jardine G0FDV

PW Publishing Ltd., Tayfield House, 38 Poole Road, Westbourne, Bournemouth BH4 9DW • E-mail: dj@ukpc.net

Transistors

NEW SERIES PART 1

David Jardine G0FDV addresses what you've always wanted to know about transistors but were afraid to ask!

The great thing about electronics and amateur radio is that we learn something new every day (well almost) which makes it a fascinating hobby. Transistors are everywhere, literally. Probably there are millions, even billions in your home and car – in the computer, TV, radio, washing machine, TV remote, mobile phone, DVD player, tablet, watch, ordinary house phone audio system, satnav, calculator and so on.

Recently I walked around my house and office and the number of digital and other electronic powered devices and hence transistors was absolutely astounding. I run a small computerised PR, advertising, web design and brochure business.

Transistors are probably the single most incredible items to have totally revolutionised all life on our planet today in just a few short years. They are used extensively in the aerospace industry, defence, printing, IT, banking, government, and the travel industry as just a few examples. They are truly ubiquitous, affecting us all daily, all day and every day.

Let's Get Started

A transistor is made up of the words transferable and resistor, hence tran-sistor – in other words resistance to a signal is changed by application of another, often smaller, signal.

The study of transistors is a huge subject, encompassing chemistry, electron flow theory, magnetic fields, nuclear physics and atomic structure. Here I'm going to explain just the basics that should be of interest and help to the radio amateur in practical terms.

Firstly, there is much debate and discussion and, yes, disagreement on the facts. I have trawled the internet, read several books and articles on the subject and present my findings here. This is, however, only a really very basic and superficial view of the subject.

Most electronics enthusiasts and radio amateurs know what a transistor can do but few, including myself (until recently), have worked out the whys and wherefores of how they actually work, how they amplify a signal and so on.

Wikipedia's definition is good and states the following: *"A transistor is a semiconductor device used to amplify and/or switch electronic signals and electrical power. It is composed of semiconductor material with at least three terminals for connection to an external circuit. A voltage or current applied to one pair of the transistor's terminals changes the current through another pair of terminals. Because the controlled (output) power can be higher than the controlling (input) power, a transistor can amplify a signal. Today, some transistors are packaged individually but many more are found embedded in integrated circuits.*

"The transistor is the fundamental building block of modern electronic devices and is ubiquitous in modern electronic systems. Following its development in 1947 by American physicists ***John Bardean****,* ***Walter Brattain*** *and* ***William Shockley****, the transistor revolutionised the field of electronics and paved the way for smaller and cheaper radios, calculators and computers among other things. The transistor is on the list of IEEE milestones in electronics and the inventors were jointly awarded the 1956 Nobel Prize in Physics for their achievement".*

Shockley also went on to be involved in the commercial manufacture of transistors. Here I will concentrate on the bipolar junction transistor (BJT), which is constructed from semiconductor material.

Transistors are often referred to as semiconductors because they comprise material that only 'partially' conducts electricity. In this article I will concentrate on silicon constructed transistors (I will ignore germanium, used in many early devices and still found occasionally for particular applications, although some of the principles are the same.)

Let's first of all, consider silicon, the most popular material used in semiconductors. On its own it is a pretty useless conductor of electricity. In fact, from an electrical viewpoint it is useless and can be considered as an insulator or certainly an extremely bad conductor. To enable silicon material to partially conduct (and therefore become a semi-conductor), it has to be doped by a special process.

The silicon atom comprises a central nucleus with 14 positively charged protons at its centre, surrounded by 14 electrons that orbit the nucleus in three different rings or shells. The inner ring contains two electrons, the next ring contains eight electrons and the third ring contains four electrons. These outer four electrons join up or share with the four electrons in nearby atoms so that a crystal lattice is formed made up entirely of silicon. By adding or doping the silicon with another

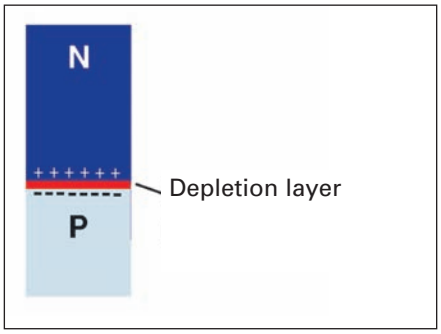

Fig. 1: A PN junction with no voltage applied.

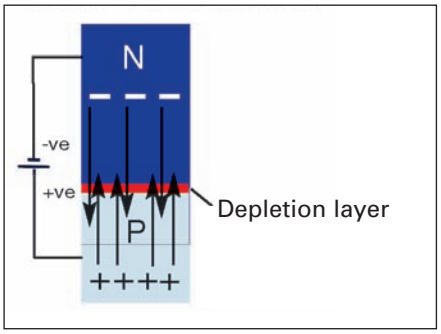

Fig. 2: A PN junction (diode) with forward bias.

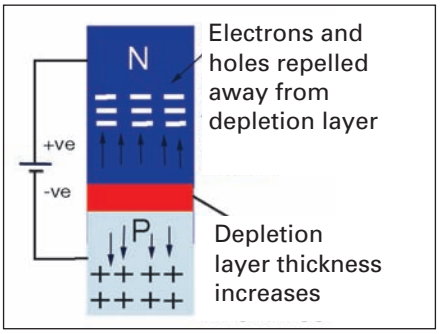

Fig. 3: A PN junction, reverse biased.

atom of, say, phosphorus or boron, the lattice takes on different properties (as explained below) and this new material is then able to partially conduct.

N- and P-Type

The two types of semiconductor materials in a basic BJT. The first is known as N-type material. In N-type doping, phosphorus or arsenic is added to the silicon in small quantities. In the phosphorus and arsenic atom there are five outer electrons. These are introduced into the silicon lattice through a process called doping. The first four electrons join with the four in the outer ring of the silicon atom to form a complete shell (an atom can only have a maximum of two electrons in its first shell and eight in each of the subsequent ones). The fifth electron is free to move around, allowing an electric current to pass through the doped material when a voltage is applied. It takes only a very small quantity of the impurity to create enough free electrons to allow an electric current to flow through the silicon. N-type silicon now becomes a conductor. The lattice appears to be negative because there are now 'surplus' electrons floating around. Electrons have a negative charge, hence the name N-type.

The other material we need to consider is P-type. In P-type doping, boron or gallium is the dopant. Boron and gallium have three electrons in their outer shell. If this material is mixed into the silicon lattice, so-called 'holes' form in the lattice because the outer shell is now incomplete, with only seven electrons in total. The absence of an electron to complete the shell creates the effect of a net positive charge, hence the name P-type. Holes can, in effect, conduct current. A hole happily accepts an electron from a neighbour to fill the empty space, thus moving the hole over to the atom that has given up its electron. As a result, P-type silicon is also able to conduct an electric current.

A minute amount of either N-type or P-type doping turns a silicon crystal from a good insulator into a viable (but not great) conductor, hence the name semiconductor.

How Transistors Work

To more easily understand what goes on in a transistor, let's start by considering the workings of a simple diode because we will understand more clearly what controls the transistor. In fact, it's all down to what goes on in the depletion layer – the area where the N-type dosed silicon meets the P-type dosed silicon. This is where interesting things start to happen.

When we study a junction where N-type and P-type materials come into contact as in a diode, **Fig. 1**, immediately some electrons in the N type (negative charges) jump towards the junction and, conversely, holes are attracted towards the junction by the negative charge on the opposite side.

What happens is that the holes and electrons combine, creating a small area of non-conducting silicon known as the depletion layer. What happens is that the 'surplus' electrons on one side of the junction fill the holes on the other, restoring the status quo. The silicon now reverts to its undoped state and is no longer able to conduct. The junction, therefore, becomes a barrier to current flow.

The only way to get current to flow is to create a potential difference across the device, attaching the positive supply to the P-type material and the negative to the N-type because like poles repel, **Fig. 2**. The result is that charge carriers (electrons and holes) will be forced towards the junction and current flow will occur. For silicon, the voltage must be in the region of 0.5 to 0.7V to start the flow.

If we therefore connect a supply (a battery, for example) of at least, say, 0.7V across the semiconductor junction as shown in Fig. 2 with the negative terminal to the N-type and positive terminal to the P-type, electrons will now happily cross the junction from N- to P-type and combine with the holes in the P-type due to the fact that 'like poles repel' and 'unlike polarities attract' each other. In the same way, the positive holes can be considered as crossing to combine with the negative electrons in the N-type material. A current flow or charge flow is therefore set in motion. This condition is known as 'forward bias'. It is vital to remember in this case that when N-type material is connected directly to P-type material, the depletion layer, mentioned earlier, creates a barrier potential adjacent to the two materials that will actually oppose the flow of current unless sufficient forward voltage exists across the layer to overcome this barrier potential. Without this, the diode at the point of the depletion layer (which is typically only .001mm thick) returns to its original silicon state with no charge carriers.

When the polarity of the power source is reversed, **Fig. 3**, so that the negative terminal of the voltage source is connected to the P-type material, the holes are attracted towards the negative polarity of the supply and the electrons in the N-type are attracted to the positive terminal in like fashion. In this case, therefore, the depletion layer is widened considerably and no current will flow because there are no longer any charge carriers in the depletion layer. This is known as the 'reverse bias' configuration. It needs some study and thinking about it to fully grasp what is happening.

In a diode, by studying the above, we can see that when an alternating current is applied, current flow is allowed or permitted in just one direction. It is the thickness or strength of the depletion layer that is controlling things. The same principle applies to transistors, which have three layers of material in a PNP or NPN configuration, effectively two diodes back to back, one in forward bias configuration and the other in reverse bias. I'll explain how these work next month.

PW Publishing Ltd., Tayfield House, 38 Poole Road, Westbourne, Bournemouth BH4 9DW • E-mail: dj@ukpc.net

Transistors

NEW SERIES PART 2

David Jardine G0FDV continues his explanation of the workings of transistors.

Forgive me for oversimplifying the complex theories of electron flow that involve a more professional academic understanding. Here I want to explain what is actually going on inside a transistor in more basic terms. Detailed theories of electron flow are beyond the scope of this article and I would recommend perhaps the *RSGB Handbook* or similar for a deeper study. I have also given some online references at the end of the article.

The diagram, **Fig. 1**, (forward biased) shows a sandwich of NPN material as in a bipolar junction transistor (BJT), as we are discussing here. The N-type on the bottom we will call the emitter, the P-type in the middle we will call the base and the N-type on the top we will call the collector.

When we bring out our battery again as in the case of the diode (see Part 1 in last month's issue) and connect its negative terminal to the N-type emitter, a small positive voltage to the P-type centre (base) section and connect the positive terminal of the battery to the N-type top (collector) what is going to happen?

Electrons will readily flow across the junction from emitter to base and onwards to the collector and holes will flow from the base to the emitter (although, as I explain below, it's not actually the holes that flow but, rather, electrons in the opposite direction!). In other words current flow is activated as I described in the context of diodes. Now remember that the centre P-type section is very lightly doped, is extremely thin and contains very few holes compared with the electrons in the two N-type sections.

What is Happening

What is happening is that this small base voltage sets in motion positive charge carriers – more holes go into the base as electrons are withdrawn. In effect, holes (charge carriers) attract lots of negative electrons across the junction. Because the junction is so thin, most of these electrons are attracted to the positive polarity at the collector and a large current will flow, varying in exactly the same proportion as changes that are made in the base current or voltage. We can see, then, that with careful adjustments of the base bias (the small positive voltage on the base) and/or the application of an alternating signal at the base, we can control a larger current flowing between collector and emitter. In other words, we have effectively amplified the input signal (current).

Conversely, **Fig. 2**, if the positive voltage at the base is withdrawn, the depletion layer is widened and the current will reduce or stop completely. At the point at which current flow ceases, the transistor is well and truly 'turned off'.

I need to explain here that 'holes' represent a deficiency of electrons, so a flow of 'holes' is actually a flow of electrons in the opposite direction, filling each hole in turn as they cross the junction. Both 'holes' in one direction and electrons in the other can, in effect, conduct current. It is simply a movement of charges that is taking place.

If the base voltage is too low (lower than 0.5V above emitter (as in Fig. 2)), bound electrons will accumulate in the P-region and no current will flow across the junction. So the depletion layer is in control here and becomes wider. Too low a voltage and the transistor will be switched off. Once the voltage exceeds 0.5V, in the case of silicon, the transistor will start to conduct.

Another way of looking at this complex situation is that we are, in a sense, 'sucking out' electrons from the base and creating lots of holes that encourage electron flow across the entire semiconductor when the transistor is forward biased. In fact, if one electron is drawn out of the base by the positive voltage at the base, something like 50 will jump across from emitter to base and on to the collector! Hence the small change in base current is superimposed on the much larger operating current across the entire transistor. It is the positive potential at the collector that is attracting these electrons across the high impedance depletion layer between the base and collector.

An Example

To take an example, let's suppose the base current I_b is of the order 5μA when the collector current I_c is 1mA. Because the emitter base junction is forward biased, a very small increase in the base voltage, say from 500mV to 520mV, may be sufficient to cause an increase in the base current from 5μA to 10μA and an increase in collector current from 1mA to, say, 2mA.

As I said above, typically 50 electrons cross over from emitter to collector before one falls into a hole. This is because the base region is extremely thin, is far less heavily doped than the emitter's N-type

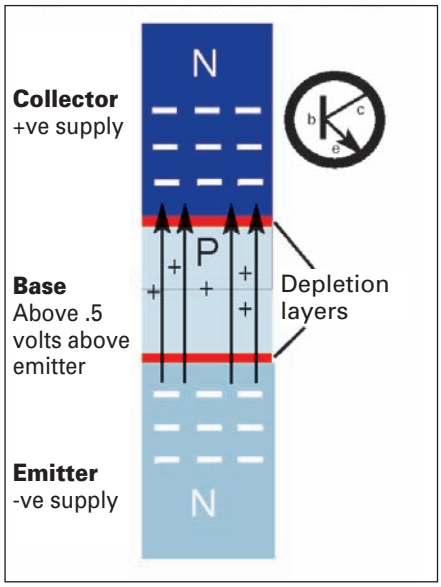

Fig. 1: A forward-biased bipolar junction transistor.

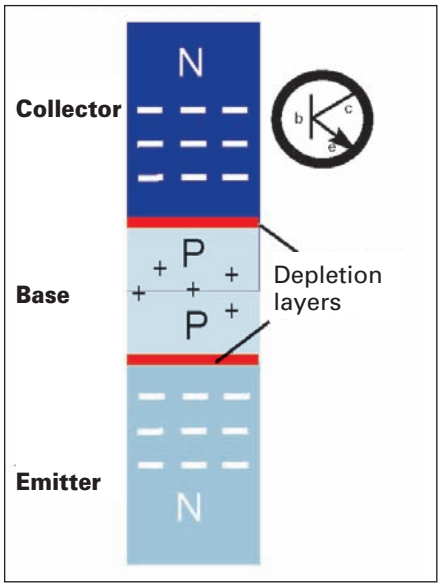

Fig. 2: The transistor of Fig. 1 but biased such as to cut off current flow.

material and the thickness of the depletion layer has been reduced so that current flows across the junction. Therefore a small base current is controlling a much larger collector current.

The transistor therefore has the appearance and, indeed, the virtues of an amplifier but in a sense is merely superimposing the base current changes on to the much larger collector current.

Semiconductor theory states that when a signal is applied between the base and emitter of a BJT circuit, an increase in voltage on the base leads to an increase in current through the base and collector, causing the voltage at the collector to decrease, hence a 180° phase shift between input and output signals.

The emitter/base junction input has a lower impedance compared with the high impedance of the reverse biased base/collector junction in the case of a common emitter BJT configuration.

Transistor Characteristics

When looking at the specifications of transistors, you will see several terms being referred to. These include:

Current amplification, Alpha is a measure the amount of current in the emitter circuit that finds its way into the collector circuit. That is to say, the change in collector current divided by change in emitter circuit. It is usually nearly unity and seldom less than 0.98 (the remaining current is that which flows through the base connection).

Current gain, Beta – important to constructors – is the change in collector current divided by change in base current that has caused it. This usually lies usually between 50 and 500 – handy for that RF preamplifier!

Transition frequency is the highest frequency at which the transistor can amplify. In practice, the amplification will fall off with frequency.

The values of each of the above will depend on the design and construction of the transistor. The junctions of a power transistor, for example, might have a large area to handle the large currents needed. This in turn will cause the junctions to have a higher capacitance than other types, resulting in a fall-off of gain at lower frequencies than would otherwise be the case. And just try to imagine the difference between a transistor that is part of a modern microprocessor with maybe a billion transistors on one chip and, conversely, a device designed to handle large currents in a power supply or solid-state linear amplifier running several hundred watts. Transistor design has certainly come a long way and, of course, nowadays includes other configurations such as field effect transistors (FETs) that work in a slightly different manner to that which I have described although the underlying principles are the same – transistor fundamentals have remained similar for several decades now although technology has enabled them to be built to more exacting standards in terms of size, power handling and so on.

If you want to pursue the subject further, there is lots of information online and in the textbooks, some of which can be very technical and very mathematical! However, a search via your favourite search engine is a good place to start. Below, for example, are just three sites that might prove interesting, while there are also plenty of explanatory videos on YouTube.

www.explainthatstuff.com/howtransistorswork.html
www.build-electronic-circuits.com/how-transistors-work
https://learn.sparkfun.com/tutorials/transistors

Help support the...

RADIO COMMUNICATIONS FOUNDATION
Registered Charity number 1100694

- The Foundation is a Registered Charity that exists to help young people get into radio and electronic engineering and to promote radio technology. It has helped fund the RSGB's educational outreach programme, provided finance for projects put forward by clubs or educational institutions, awarded scholarships to youngsters who demonstrate a commitment to radio and engineering and handled bequests for those who want their enthusiasm for radio remembered by future generations.
- What can you do? Your donation, regular payment or bequest in a Will is going to help the Charity continue its good work to benefit others in the hobby. Support us so that we can support amateur radio and everyone who is interested and active in the hobby.
- Please visit our website

www.commsfoundation.org

to find out all about us and follow the link to Donations.

RCF – looking to the future